DOCUMENT RESUME

ED 357 954 SE 053 119

AUTHOR Helgeson, Stanley L.
TITLE Trends and Issues in Science Education.
PUB DATE Jun 92
NOTE 36p.; Keynote address presented to the International Symposium on Science Education, (Taipei, Taiwan, June 12, 1992).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Attitude Measures; Computer Assisted Instruction; *Computer Assisted Testing; Computer Uses in Education; *Curriculum Development; *Educational Trends; Elementary Secondary Education; Evaluation Methods; Integrated Curriculum; International Educational Exchange; Microcomputers; Program Descriptions; *Science Curriculum; *Science Education; Science Instruction; Standardized Tests; Student Attitudes; Test Construction; Test Validity
IDENTIFIERS *Alternative Assessment; Earth Systems Education; *Educational Issues; Project 2061 (AAAS); Scope Sequence and Coordination

ABSTRACT

This keynote address reported on the current state of science education in three areas: curriculum development, attitudes toward science, and assessment in science. Three major projects in science curriculum reform were the focus of the section on curriculum development: (1) Project 2061; (2) Scope, Sequence, and Coordination; and (3) Earth Systems Education. Discussions of the projects included background information, project objectives, and a project description. The section on attitudes toward science discussed problems related to assessing attitudes, and provided guidelines for instrument development and establishing test validity. The section on assessment in science education reported recent studies and projects related to standardized tests, computer applications in assessment, and alternative assessment. A list of 55 references is included. (MDH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED357954

Trends and Issues in Science Education

Stanley L. Helgeson
The Ohio State University
Columbus, Ohio 43210
U.S.A.

Keynote address presented to the
International Symposium on Science Education
National Taiwan Normal University
Taipei, Taiwan
Republic of China

June 12, 1992

SE053119

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Stanley L. Helgeson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Trends and Issues in Science Education

Stanley L. Helgeson
The Ohio State University
Columbus, Ohio 43210
U.S.A.

Activity abounds in science education today. Concern for student achievement in science drives curriculum modification and development, drives infusion of new technologies into instruction, drives changes in assessment techniques, and increased concern for student attitudes toward science. These activities are not limited to one or a few countries. Virtually every country reports devoting attention to similar concerns. It is beyond the scope of this paper to review all areas of activity. Rather, a few will be highlighted as indicators of current endeavors in the field, namely, curriculum development, attitudes toward science, and assessment in science.

Curriculum Development

Three major efforts in science curriculum reform and development will be reviewed: Project 2061; Scope, Sequence and Coordination; and Earth Systems Education.

Project 2061

Project 2061, an activity of the American Association for the Advancement of Science (AAAS), originated in 1985 to help bring about the reform of education in science, mathematics, and technology. The Project was initiated when Comet Halley happened to be near the Earth. That event led to the Project's name. It was realized that children who would live to see the comet's return in 2061, a human lifetime in the future, would soon be starting their school years, hence, Project 2061.

Recognizing scientific literacy as a national goal, three questions provided the central purpose of the Project: What is the substance of scientific literacy? Who should be expected to acquire the requisite knowledge and skills? And how can scientific literacy be achieved nationwide? (*Science for All Americans: Summary*, 1989:1).

Project 2061 is designed in three phases. Phase I, drawing upon five scientific panels and a wide array of consultants (scientists, engineers, mathematicians, historians, and educators), focused on the substance of scientific literacy. "The purpose of Phase I was to establish a conceptual base for reform by spelling out the knowledge, skills, and attitudes all students should acquire as a consequence of their total school experience from kindergarten though high school" (*Science for All Americans: Summary*, 1989:4). Project 2061 defines science education to include all of the natural, physical, social, and behavioral sciences, mathematics, engineering, and their interrelationships. As the major products of Phase I, six reports have been published. The overview report is *Science for All Americans*, written by AAAS Project 2061 in consultation with the National Council on Science and Technology Education. Five reports from scientific panels were also produced:

Biological and Health Sciences: Report of the Project 2061 Phase I Biological and Health Sciences Panel, by Mary Clark.

Mathematics: Report of the Project 2061 Phase I Mathematics Panel, by David Blackwell and Leon Henkin.

Physical and Information Sciences and Engineering: Report of the Project 2061 Phase I Physical and Information Sciences and Engineering Panel, by George Bugliarello.

Social and Behavioral Sciences: Report of the Project 2061 Phase I Social and Behavioral Sciences Panel, by Mortimer Appley and Winifred B. Maher.

Technology: Report of the Project 2061 Phase I Technology Panel, by James R. Johnson.

The basic dimensions of scientific literacy are set forth by the national council's recommendations as follows:

- Being familiar with the natural world and recognizing both its diversity and its unity

- Understanding key concepts and principles of science

- Being aware of some of the important ways in which science, mathematics, and technology depend upon one another

- Knowing that science, mathematics, and technology are human enterprises and what that implies about their strengths and limitations

- Having a capacity for scientific ways of thinking

- Using scientific knowledge and ways of thinking for individual and social purposes (*Science for All Americans: Summary*, 1989:4)

The recommendations cover a wide range of topics. Many of these topics are already included in school curricula (e.g., the structure of matter, the basic functions of cells, prevention of disease, communications technology, and different uses of numbers). However, the intended treatment of these topics differs from the traditional in two ways. First, the boundaries between traditional subject matter categories are softened and connections are emphasized. Second, the amount of detail that students are expected to retain is considerably less than in traditional science, mathematics, and technology courses. "Ideas and thinking skills are emphasized at the expense of specialized vocabulary and memorized procedures" (p. 5). A fundamental premise of Project 2061 is that schools do not need to teach more, but to teach less so that it can be taught better. In other words, less is more. Some topics not usually included in school curricula are also recommended. Among these are the nature of the scientific enterprise, and how science, mathematics, and technology relate to one another and to the social system in general. The recommendations also call for some knowledge of the most important episodes in the history of science and technology, and of the major conceptual themes that run through almost all scientific thinking. The council's recommendations can be summarized in four general categories: The Scientific Endeavor, Scientific Views of the World, Perspectives on Science, and Scientific Habits of Mind (p.5).

Phase II of Project 2061--Redesigning the Educational System--is currently in progress. This Phase involves teams of educators and scientists in transforming the knowledge, skills, and attitude outcomes specified in *Science for All Americans* into several alternative curriculum models for use by school districts and states.

The Project is also, during this phase, drawing up blueprints for reform related to teacher preparation, materials and technologies for teaching, testing, equity, and other school issues (Update, 1992:7). Phase II is thus developing the links between Phase I and Phase III. These links include (1) product development--developing the intellectual tools needed to reform K-12 science education; (2) outreach--increasing support for change by building alliances; (3) resources--fostering the production of instructional materials needed once implementation begins; and (4) launching reform--implementing Project 2061 at a limited number of sites (Update, 1992:11). Of the specified tasks, developing reform tools has the highest priority. Four types of tools are under development: Benchmarks for Scientific Literacy, Alternative Curriculum Models, Resource Database, and Blueprints for Reform.

Six research and development sites have been selected to represent, collectively, the demographic characteristics of school districts in the U.S. At each location a team of 25 educators, cutting across grade levels and disciplines, was assembled. Each team included 5 elementary teachers, 5 middle school teachers, 10 high school teachers, 3 principals, and 2 curriculum specialists.

The teams include teachers of different subjects and disciplines: arithmetic, algebra, geometry, and calculus; general science, biology, earth and space science, chemistry, and physics; technology, home economics, and vocational education; social studies and history; language arts; and elementary teachers who deal with many areas (Update, 1992: 13). The teachers work with a range of students including average students, students with learning problems, talented students, motivated and unmotivated students, and students with the whole range of home and community circumstances. The teachers were provided with up to 40 days of release time per year, computers at home and school, a dedicated work place in each school district, telecommunications links to each other and Project staff, consultants, and a budget for materials.

Each team approached its work in a different way. However, they all devoted time to analyzing learning patterns among children, to the achievement of particular learning goals, to collecting ideas from other teachers and administrators, and to exploring new ways to configure the learning experience.

"The first step for the teams in developing curriculum models was to determine the progression of understanding by which students might eventually arrive at the learning outcomes in *Science for All Americans (SFAA)*" (Update, 1992:14). Thus, the task was to determine what components of each outcome younger students should have in order to understand new material. This technique is referred to as backmapping. Each major concept in *SFAA* was mapped backwards to specify the preceding concepts needed to make sense of the new. Each concept was then placed at a rough grade level according to when students would be best able to learn it. As these backmaps were produced, ideas emerged for activities and learning experiences that could serve multiple understandings at each grade level. A particular concept might serve as a prerequisite for two or three other learning outcomes. The maps thus became interlaced to produce broad patterns of conceptual growth. These interconnected maps then provided teams with the context and organization of the entire curriculum (Update, 1992:15)

The learning outcomes specified in *SFAA* serve as a base for establishing benchmarks for scientific literacy. These benchmarks are expressions of learning outcomes in greater detail and at several grade levels. These benchmarks, then, are standards. Standards are viewed by Project 2061 as indicators of relative levels of achievement, or achievement norms approved by professions. When these benchmarks are established they will provide schools with another curriculum design tool to be used in conjunction with *SFAA*. The U.S. Secretary of Education has asked the National Research Council of the National Academy of Sciences to orchestrate the creation of national standards in science education. The AAAS, the National Science Teachers Association, and other groups will contribute to this effort. Both *Science for All Americans* and the benchmarks currently being developed by Project 2061 will be taken into consideration.

The second major tool for curriculum reform being produced by Project 2061 is a series of curriculum models. "A curriculum model, in the Project 2061 scheme of things, is a description of a possible curriculum with enough detail to enable educators to create an actual curriculum having the properties of the model" (Update, 1992:20). It is intended that the model will also influence the development of new learning materials and new teacher education programs. The model, then, should specify the content domain covered, the students to be served, and the grades spanned. It should also indicate the

intended learning goals, provide a rationale for a curriculum design, and describe the kinds of learning experiences that the students will have and when they will have them. Finally, the model should specify the conditions necessary for proper functioning. To summarize--Domain, Goals, Design, and Conditions. The model does not, however, include course outlines, lesson plans, materials, or a precise timetable. In contrast to a model, an actual curriculum contains the detail necessary to schedule students and carry out day to day instruction (Update, 1992:20-21).

The six school based teams are all contributing to the design of curriculum models. The models being developed differ from each other with what the teams have designated "conceptual" distinctions. Four such conceptually different models are currently evolving. A model emphasizing How The World Works focuses on explaining natural phenomena, objects, and processes of interest to students. As the students mature, the explanations are increasingly based on scientific and engineering principles and quantitative thinking. An Inquiry model would cover much of the same content but with more emphasis on science as a way of knowing. This model would emphasize science as a social and cultural endeavor as much as an individual and creative activity. A Design model would concentrate more on engineering thinking, the solution of real-world problems for which there is no ideal solution, and the understanding of the technologies that influence society. A design organized around Human Concerns would emphasize interdisciplinary studies and involve the arts and humanities as well as science. All the models will be designed to meet the 2061 benchmarks and the *SFAA* learning outcomes. All will use diverse teaching approaches--inquiry and design projects, seminars, independent study, case study, team learning and teaching by students. And all will use various print, electronic, and multi-media.

The third type of tool being developed for Project 2061 is a resource data base. School districts wanting to develop complete curricula based on Project 2061 models will need information on a variety topics such as curriculum design, research on chid development, feedback on implementation, etc. The Project is therefore developing a computerized data base that will list the most important relevant information available for teaching science, mathematics, and technology. The listings will include print, film, video, computer disk, and multimedia. The data base will be periodically revised and updated.

Finally, taking a systems approach to curriculum reform, Phase II of Project 2061 will include developing blueprints for reform. Each blueprint will indicate current theories and conditions, the requirements of the curriculum models, likely obstacles to implementation, and practical recommendations on how to achieve reform (Update, 1992:24). Areas to be included in the blueprints include teacher education, assessment, materials and technology, curriculum connections school organization, parents and the community, higher education, business and industry, educational research, equity, and educational policy.

Phase III will be a widespread collaborative effort to use the resources of Phases .I and II to bring about educational reform. This segment of the Project is expected to take a decade or more.

Scope, Sequence, and Coordination

The Scope, Sequence, and Coordination (SS&C) project was originated by Bill Aldridge, Executive Director of the National Science Teachers Association (NSTA). Aldridge derived the idea for a major reform in the structure of science education as a result of considering problems existing in secondary science education in the U.S. Students view the currently structured, textbook-driven science subjects as difficult, boring, and not relevant to their lives. Many student opt out of science as soon as possible. Most students take biology in ninth or tenth grade and over half of them take no science beyond tenth grade. About 40% of high school students take a course in chemistry and only 19% take a course in physics (Aldridge, 1992).

To counter these problems, the SS&C project calls for the elimination of tracking of students, recommends that all students study science every year for six years, and "advocates the study of science as carefully sequenced, well-coordinated instruction in physics, chemistry, biology, and earth/space science" (Aldridge, 1989b:1). Aldridge had conducted and analysis of science education in several countries and was struck by some obvious differences between science education in the U.S. and other countries. For example, in virtually all of the industrialized nations of the world, students study science in several subjects over several years. Only in the U.S. do we use the "layer cake" of biology, chemistry, and physics at the high school level. Thus, the project calls for "spacing" the study of each science out over several years, rather than the

"layer cake" curriculum in which science is taught in year-long, discrete and compressed separate disciplines.

There are three major components to the rationale underlying SS&C. Research indicates that students learn and retain new material better if they study it in spaced intervals rather than all at once. The students can then revisit a concept or idea at successively higher levels. The SS&C project also calls for sequencing of instruction, taking into account how children learn. Thus, they should encounter the new concept or idea first in direct, hands-on experience. Only after experience with a phenomenon should it be given a name. The third component of NSTA's Scope, Sequence, and Coordination Project is the coordination of science concepts and topics. Biology, earth/space science, chemistry, and physics all have certain features and processes in common. Coordination among these disciplines would lead to awareness of the interdependence of the sciences and how they fit together as a part of the larger body of knowledge. Presenting a concept in two or three different subjects and contexts helps to establish it more firmly in the student's mind (Aldridge, 1989b).

What is proposed, then, is a restructuring of the way in which the sciences are presented, as illustrated in the following table:

Proposed Example of a Revised Science Curriculum For Grades 7 Through 12 in the United States

	Grade Level						
	7	8	9	10	11	12	Total Time Spent
Hours Per Week By Subject							
Biology	1	2	2	3	1	1	360
Chemistry	1	1	2	2	3	2	396
Physics	2	2	1	1	2	3	396
Earth/Space Science	3	2	2	1	1	1	360
Total Hours Per Week	7	7	7	7	7	7	
Emphasis	descriptive; phenonenological		empirical; semi-quantitative		theoretical abstract		

From Aldridge, 1989a, p. 7

The SS&C approach is designed to follow Piaget's ideas about how students construct knowledge. The research of Piaget and others suggests that concrete experiences with science phenomena should precede terminology. In the same vein, concepts should be derived from experience with phenomena in different contexts. After the concepts are established, they can be symbolized and those symbols related to each other. The more complex relationships would be constructed over time. A synthesis of the various research led to the conclusion that a framework should be built around three fundamental questions: What do we mean? How do we know? and Why do we believe? (Aldridge, 1992:14).

The first question implies recognition of the fact that the natural world requires language to explain it. However, it is not sufficient to just use names and terms. These are arbitrary and by themselves have no meaning. A further implication of the question is that students must first have experience with the phenomenon that a term represents, in its various contexts, before the term gains meaning. The second question, How do we know?, concerns facts of science. These facts may take many forms. They may be measurements, perhaps quite complicated measurements. The important consideration is that students know how the measurement was made. Understanding how we know leads to more confidence in the fact being correct and in better understanding of its limitations. The same holds true for empirical laws stated as facts. It is important that students understand how we know that an empirical law is valid. "Unsupported assertions of such laws is the worst kind of teaching!" (Aldridge, 1992:15). The third question, Why do we believe?, involves the reasons why we believe certain theories. The theories should not be studied until the observations, phenomena, facts, concepts, and empirical laws that the theories attempt to encompass are understood. Thus, a progression is implied. Qualitative relationships should be studied first, then come measurements and empirical relationships, and, finally theories should be constructed from which new predictions can be made. In this progression, science teachers would go from direct experiences with phenomena to terms and concepts and then to laws, principles, and theories (p. 16).

To aid school districts in implementing Scope, Sequence, and Coordination, NSTA appointed four curriculum committees to identify science topics to be taught in each discipline at each grade level. The

committee members were experts in the discipline areas and included university professors, classroom teachers, curriculum developers, and textbook authors. The committees examined a variety of sources including the goals identified by Project 2061, the major textbook series, science trade books, and a number of curriculum guides. An initial list of topics was formulated for all secondary grades, following which the committees concentrated on the seventh grade, since the first trials are to be implemented at this level. Four major themes are to be used to unify the science: Observation and Measurement; Properties and Structure; Changes Over Time and Transformations; and Periodic and Cyclic Phenomena.

Another major aspect of the project is the development of an interactive compact disc approach to assessment. A prototype disc guides students as they carry out tasks with real objects and phenomena. The system tracks the student's responses so that patterns of preconceptions, strategies, and techniques for solving problems can be identified. The system will present items at several levels of complexity. It is intended that all students can succeed at the lowest level of the item; as the item gets progressively more difficult more students are exited from the system (temporarily, they can later return). The items are designed so that only the most talented students can respond at the top level. In most cases, no student fails the item and no student fully masters it. In addition to the performance assessment, the system will test for cognitive knowledge. There will be parallel items to determine how well students understand the concepts. When the prototype disc is completed and has been tested, a series of science test items using the interactive compacxt disc will be devloped for grades 7-12.

Projects are currently underway in California, North Carolina, Iowa, Puerto Rico, Texas and Alaska.

Earth Systems Education

The Earth Systems Education project is based on the theory that there is a need to integrate the sciences in teaching science, K-12. The philosophy is consistent with that of Project 2061 and Scope, Sequence, and Coordination, but with a somewhat different orientation. Earth Systems Education focuses the study of science on the planet Earth (Mayer, 1992). Guided by the Framework for Earth Systems Education (Mayer, 1991), developed by educators and

scientists working together, the program integrates all scientific disciplines to teach science that is practical and relevant to students.

Over the past two decades there have been significant advances in the understanding of planet Earth, in part due to the use of satellites in data collection and supercomputers for data processing. These advances prompted the organization of a conference of geoscientists in April, 1988, to consider the implications of these new understandings for science curriculum renewal. The 40 scientists and educators developed a framework of four goals and ten concepts about planet Earth that they believed every citizen should understand (Mayer, 1991).

In 1990, The Ohio State University, under a grant from the National Science Foundation, began developing leadership teams in Earth Systems Education--the Program for Leadership in Earth Systems Education (PLESE). This program was designed to infuse more content relating the modern understanding of planet Earth into the nation's K-12 science curricula (Mayer, 1992). In preparation for the PLESE program, a planning committee developed a conceptual framework to guide the program. This framework now provides a basis for PLESE teams to construct resource guides and to select teaching materials for use in infusing Earth systems information into the science curriculum in their local districts.

The PLESE planning committee purposely arranged the understandings into a sequence. The first emphasizes the aesthetic values of the Earth as interpreted in art and music. Focusing on students' feelings toward the Earth systems, the way in which they and others experience and interpret their feelings, draws the students into a systematic study of the planet. An aesthetic appreciation of the planet then leads the students naturally into a concern for the proper stewardship of its resources, the second understanding of the framework. Developing a concern for the aesthetic and economic resources of our planet leads to a desire to understand how the various subsystems work and how we study these subsystems, which is the substance of the next four understandings. In learning how the subsystems function, the students must master basic physics, chemistry, and biology concepts. The last understanding deals with careers and avocations in science, thus bringing the focus back to the immediate concerns and interests of the students (Mayer, 1992:2).

The trend toward integration of the K-12 curriculum, especially driven by Project 2061, is recognized and encouraged by the Earth Systems Education project. It seems most natural, then, to develop a science curriculum using the subject of all science investigations--planet Earth--as the unifying theme. Any physical, biological, or chemical process that citizens must understand to be scientifically literate can be taught in the context of its Earth subsystem. This is the basic idea that has guided those involved in developing Earth Systems Education (Mayer, 1992:2).

There are several projects underway to test various aspects of Earth Systems Education. The major one is the PLESE program which includes participants from all 50 states. Summer workshops of three weeks duration are offered by the University of Northern Colorado and The Ohio State University for three-member teams representing elementary, middle, and high schools. The teams develop resource guides for infusing Earth Systems concepts throughout existing K-12 curricula. After returning to their school systems, each team conducts at least two Earth Systems Education workshops at the state and local level. During the last year of the project the guides that have been developed will be edited and compiled into comprehensive Earth Systems Resource Guides for each of the grade levels, and distributed nationally (Mayer, 1992).

A second project is the development of an integrated Biological and Earth Systems (BES) science sequence for the high schools in the Worthington (Ohio) School District. This two-year sequence replaces Earth science at the ninth grade and biology at the tenth grade. The sequence is organized around basic Earth issues such as resource supply, global climate change, and deforestation. The major instructional strategies used are collaborative learning and problem solving techniques. The teachers involved have made a commitment not to use a textbook; instead they identify a variety of readings from current literature and direct their students to these materials. The program also relies heavily on technology to support the restructuring process. A variety of data bases are used by the students for collecting data and current information. Students use word processing, spreadsheets, and data base programs for storing and analyzing data as well as simple data analysis programs. Students also use CD-ROM and video discs as sources of information (Mayer, 1992).

A third effort is also underway in central Ohio. Ten school systems have each designated a team of three to five middle school teachers to participate in a collaborative university-school project to consider the implications of Earth Systems Education philosophy and methods for restructuring their middle school science curricula. The teams at this time are beginning to draft syllabi for their respective school districts. Similar activities are being initiated in New York and Colorado.

Assessment of Attitudes in Science Education

Interest in attitudes toward science is by no means a new phenomenon, but it currently is drawing increased attention in the overall assessment efforts at national, state, and district levels.

Assessment Problems

Several major reviews of studies related to student attitudes toward science have been conducted in the past two decades. Omerod and Duckworth (1975) summarized the results and implications of more than 500 attitude studies. Gardner (1975) evaluated results of studies and instruments used and noted that it was possible to distinguish two broad categories: attitudes toward science and scientific attitudes (p. 1). Gauld and Hukins (1980), in a review of scientific attitudes, identified as a major problem the lack of agreement about the meanings to be attributed to various terms that are used. Munby (1983) examined the problems of assessment and instrumentation through review of more than 50 instruments used to assess attitudes. Schibeci (1984) updated the research on attitude toward science and presented general conclusions and issues from more than 200 studies. Two problems appear consistently: lack of conceptual clarity in defining attitude toward science, and difficulty with instruments used to assess attitudes (Krynowsky, 1988). Drawing on the writing of Blosser (1984), Gardner (1975), and Munby (1983), Germann (1988) states:

> First, the construct of attitude has been vague, inconsistent, and ambiguous. Second, research has often been conducted without a theoretical model of the relationship of attitude with other variables. Third, the attitude instruments themselves are judged to be immature and inadequate (p. 689)

Guidelines for Instrument Development

In their review of assessment instruments in science, Mayer and Richmond found that few instruments were submitted to repeated use and continual refinement. They also found extensive duplication of efforts, especially in the area of assessment of attitudes toward science and scientists. Since numerous attitude assssment instruments are available, they contend that efforts should be directed toward the revision or refinement of these instruments. Based on the work of Gardner (1975), they suggest guidelines for the development and refinement of such instruments:

1) The specification of a clear theoretical construct to underlie the instrument. This construct should then guide the selection and development of items.

2) Avoidance of confusion between different theoretical constructs. If more than one is to be included in a single instrument, each should be identified and scored as a separate factor.

3) The elimination of defective items such as those that combine two or more understandings or perceptions.

4) The preliminary trial of the instrument on a population with characteristics approximating those of the population with which the instrument is to be used.

5) Provision in the design for filtering out influences of respondent knowledge about the scientific enterprise from attitudes toward it.

6) The refinement of the instrument for each factor or subscale such that a reasonable internal consistency is obtained.

7) Determination of the stability of the instrument through test-retest techniques.

8) The use of factor-analysis to empirically validate factors (p. 61).

The authors contend that a few attitude instruments developed and refined in this manner will prove more useful than the continued generation of unrelated instruments in isolated studies.

Establishing Validity of Attitude Instruments

Based on the contention that validity testing of an instrument is dependent upon there being a conceptual link between the construct being measured and another construct, Munby, Kitto and Wilson (1976) describe the application of the multitrait-multimethod model (Campbell and Fiske, 1959) to determine validity.

	Trait 1	Trait 2	Trait 3
Method 1	Test A	Test C	Test E
Method 2	Test B	Test D	Test F

The conceptual significance of the Campbell and Fiske model may be described as follows. Suppose the validity of a test reputed to measure trait 1 (Test A) by an observation instrument (Method 1) is in question. First, the measure should correlate significantly with Test B, since it measures the same trait with a different method, a written test (Method 2). This multimethod axis of the model assures us that any validity score of Test A is not an artifact of the method of measurement used. Second, so that we can be assured that A and B measure a construct which is not isolated but coheres with another construct having a defined conceptual coherence, a second trait representing the coherent and converging construct is identified, and instruments C and D are selected with the intent of having A and B scores correlate significantly with those of D and D. Third, in order that A and B are measures of a distinctive trait or construct, (that is, they don't converge with everything) it is necessary to show that they discriminate from other measures, E and F, of a construct (trait 3) conceptually understood to have no relationship to traits 1 and 2. The theoretically derived hypotheses are that measures A, B, C, and D will correlate positively and significantly, while measures E and F, while correlating

positively and significantly with each other, will have zero correlations with each of A, B, C, and D scores (p.314-315).

Following are two examples of attitude instrument development based upon both a theoretical foundation and empirical evidence. Krynowsky (1988) used the Azjen and Fishbein theory to guide the development of the Attitude Toward the Subject Science Scale (ATSS). This theory posits a theoretical relationship between attitude and behavior. This is interpreted to mean that "a student attitude toward the subject science is defined as a learned predisposition of an individual to respond in a consistently favorable or unfavorable way to performing behaviors related to the teaching/learning of the subject" (p. 579). Thus, an attitude assessment involves students evaluating the prospective performance of these behaviors. The ATSS was developed, refined, and tested for reliability and validity. Test-retest reliability coefficients of 0.82 and 0.84 were obtained. Validity was established by two techniques. First, teachers of two science classes were asked to rank students in terms of most positive attitude toward the subject science and the students' rank order was correlated with the students' rank order on ATSS scores. Spearman-rank order scores were 0.79 (n=25) and 0.65 (n=19) for the two classes. The second approach compared student ATSS scores to scores obtained on a reliable attitude toward the subject science scale, the School Science scale used in the British Columbia Assessment. The correlation of student scores was 0.70 (p. 581). It was the intent of the investigator that this example of using a theoretical/empirical basis for development could be further pursued by other science education researchers.

Drawing upon Fishbein's view that beliefs and behavioral intentions are determinants of attitude, Germann (1988) developed the Attitude Toward Science in School Assessment (ATSSA) as a measure of a unidimensional concept of attitude. The ATSSA "was to measure a single dimension of a general attitude toward science, specifically, how students feel toward science as a subject in school" (p. 694). The instrument consists of a Likert-type scale with five responses ranging from strongly agree to strongly disagree. The instrument was subjected to a series of pilot tests in which student scores were correlated with their own estimates of attitudes and with teachers' estimates, and to principal-component factor analysis. The revised instrument, consisting of 14 items, was then field tested in four studies. In all four studies, Cronbach's alpha estimates of

reliability were all greater than 0.95. All 14 items were found to load on only one factor with consistent factor loadings in all four studies. Discrimination was demonstrated by item-total correlations which ranged from 0.61 to 0.89. Attitude scores were compared from two classes, one of which was marked by poor discipline and inappropriate teacher behavior and the other by more experienced and skillful teaching. A t-test showed a significant difference between mean scores ($p = 0.001$) with the students in the more experienced class having more favorable attitudes. In a comparison of scores from two classes with equally skilled teachers a nonsignificant difference was found. While Germann notes that this evidence lacks experimental control, it speaks to the ability of the instrument to discriminate. In applying the instrument to study the relationship between attitude and achievement, he found relatively low correlations and concluded that other factors were more important in the kinds of learning measured by achievement tests. However, achievement including an evaluation of the consistency and quality of classwork as incorporated in a course grade seemed to be more strongly correlated to attitude.

From the evidence it is clear that ambiguity of terms and quality of instruments are two serious problems facing those interested in assessing attitudes to science. The lack of a theoretical base has been identified in nearly all cases as a hindrance to assessment. Further, the lack of empirical support for most of the existing instruments exacerbates the situation. On the encouraging side, techniques for establishing the validity of an attitude assessment instrument have been detailed by Munby, Kitto and Wilson (1976). And, commendable attempts have been made to establish a theoretical foundation for assessment by Krynowsky (1988) and Germann (1988). Taken in combination, the guidelines and foundation exist, it now remains to capitalize on them. However, a note of caution is in order. As Germann points out, care must be taken in determining what factors are to be examined in relation to student attitudes. Gender, family influence, home environment, self-concept, and peer pressure all appear to influence student attitudes (Gardner, 1975; Simpson and Troost, 1982).

Assessment in Science Education

Reasons for assessing student learning in science have been identified as: improvement of science instruction and programs; conveying expectations to students, parents, teachers, and

administrators; monitoring the status of individuals, classes, districts, states, and the nation; and accountability (Raizen et al., 1989, p. 10).

Standardized Tests

As has been widely reported, achievement scores on national assessments of science have not been encouraging. "Trends for 9-,13-, and 17-year-olds across five national science assessments conducted by NAEP from 1969 to 1986 reveal a pattern of initial declines followed by subsequent recovery at all three age groups. To date, however, the recoveries have not matched the declines" (Mullis and Jenkins, 1988:5).

With the call for accountability in the past decades came an increase in state-mandated assessment of student achievement. Most often this is accomplished by means of standardized tests. While testing is held by many to benefit education, the validity and value of traditional standardized testing is increasingly a subject of debate.

Because of the many negative factors attributed to the standardized tests used for much of student assessment, Herman and Golan (c.1992) conducted a study to discover more about the specific effects of standardized testing on teachers and classroom instruction. They surveyed more than 340 teachers in 48 different schools and identified definite relationships between standardized testing and the teaching and learning process. Among other things, they found that teachers felt strong pressure, especially from district administrators and the media, to improve their students' test scores. Administrators spent considerable time discussing with teachers ways to improve test scores and provided teachers with materials to support students' test taking skills. This finding is in agreement with results recently reported by Mertens (1992) in a study conducted in the state of New York.

Herman and Golan also reported that testing substantially influenced teachers' classroom planning. Teachers made sure that their instructional programs covered test objectives and many teachers looked at prior tests to assure a good match. Teachers also adjusted curricular scope and sequence based on test content and students' prior performance. Further, teachers devoted substantial time to test preparation activities, test-wiseness instruction and practice tests p. 59-60). The results of the tests were, in the teachers' minds, of uncertain meaning and of uncertain value in school improvement.

Teachers did not believe that standardized testing was helping schools to improve or that testing helped clarify school goals, provide useful feedback, or assess the most useful learning for students (p. 61-62).

The Elementary Science Program Evaluation Test (ESPET), administered to all fourth graders in New York State, was designed as an instrument to promote change. Cognizant of the danger that test scores alone can lead to unfortunate inferences, the ESPET developers added a non-mandatory component to the instrument to collect "information about student attitudes toward science, and about elements of the science program as perceived by students, administrators, teachers and parents/guardians...." (Mertens, 1992:2). These qualitative data were intended to help explain student performance on the required components of the test. Mertens chronicled the events which preceded and followed administration of the test in one relatively small, suburban district.

Mertens found, among other things, that the assumptions held by the superintendent, the principal, the teachers, and the science curriculum director were rarely clear to those involved and, instead of being shared, were, in fact, in direct conflict much of the time. The science director's attempts to improve the science program brought about changes that she and the teachers believed would improve the program, only to find that student scores on the ESPET declined. Shortly afterward, the test scores for every district in the county were published in the local paper. Resulting pressure from the administration to raise scores led to behaviors much like those described by Herman and Golan and away from the changes in the science program that had been sought by the teachers and science director. The dramatically improved science scores on the third year tests seemed to vindicate the actions taken to improve scores. Mertens states, "It is highly questionable, however, whether these higher scores can be interpreted to mean improvement in the school's science program" (p. 14). Rather than encouraging administrators to engage in a wide-ranging examination of the purpose, process, and structure of the program based on the test results, "the pressure to improve results puts educators in the position of having to rely on 'quick-fix' activities that are narrowly focused on the tests themselves. This study clearly supports that conclusion" (p. 15).

The negative attitudes of teachers toward standardized testing reported by Herman and Golan appears to be justified. Too often the

tests appear to result in comparison of schools and districts, rather than in informed science program improvement. A documented outcome of standardized testing would seem to be the teaching of test-taking. In fact, there is some suggestion that testing may actually result in less science being taught, and that of less value. All too often program evaluations are reduced to single scores, with little evidence that these scores are related to the science program, instructional approach, or student understanding of science. On the brighter side is the fact that there are attempts to include other facets in statewide assessment. Florida, Missouri, California, Texas, and New York, to cite just a few examples, all are endeavoring to include some type of performance in the assessment effort, even if the attempts are not always as successful as one might like. Finally, the advent of computer assisted assessment and performance based assessment offers the opportunity to evaluate a greater range of outcomes with greater flexibility.

Computer Applications in Assessment

In summarizing research findings on computer-based education, Waugh and Currier (1986) found that: (1) groups experiencing some kind of computer-based education attained test scores which were on average between .25 and .44 standard deviations higher than their comparison groups; (2) there was evidence favoring the use of computer-based education with academically disadvantaged students; (3) long term retention was no better for computer-based education than for other modes of instruction; (4) secondary students who experienced computer-based education had more positive attitudes toward computers than did their peers who did not experience computer-based education; and, (5) there was significantly less time required for computer-based education compared to conventional instruction. It should be noted that many of the studies summarized relied heavily on drill and practice modes of instruction. Such programs depend upon immediate feedback as a major function. While this may not fit the common perception of assessment, it is clear that it does in fact function in such a manner and that the immediate feedback may well have a positive impact on learning.

A common use of computers in assessment is to provide teachers with access to large banks of items for testing. These may range from specific topics such as medical biochemistry (Aesche and Parslow, 1988) for instructors of a given course, to a test bank

designed for state assessment (Willis, 1988), to a broad range of juried test items which teachers anywhere in the country may access and download onto their own computers (Dawson, 1987). Once the item banks are in place, the computer may then be used to devise unique combinations of test items for each student and to use the results of those tests to develop remedial learning activities for each student. In each case, the computer can administer the quizzes, grade and record the results, and provide the student with immediate feedback (Dunkleberger, 1980). Use of the computer to file test questions, assemble examinations, handle all records, produce and grade tests, and guide students to what should be done next enables testing to be done with an efficiency not possible from any teacher (Summers, 1984; Vogel, 1985; Heikkinen and Dunkleberger, 1985).

Leuba (1987) argues that machine-scored testing is appropriate, efficient, and effective in basic engineering sciences, especially in large classes. His arguments would seem to apply equally well to computer-assisted testing, particularly since computers are clearly a component of much of engineering science today. Leuba maintains that certain conditions apply. First, certain basic knowledge should be instilled in a student's "lifetime" memory, and that such knowledge should be explicitly tested for. Second, an upcoming test should be the stimulus for learning how to learn, i.e., practice in becoming proficient in new technical matter within a limited time (this is especially important given the competing demands on students' time). Third, an impending test stimulates students to sharpen their problem-solving skills, and the test should measure problem-solving skills. Fourth, testing should promote learning. The test should not only be an impetus to study, but a well designed test should reinforce learning. By using machine-scored tests, students can be presented with four times as many questions as can be handled in the same time if hand-scoring techniques are used. Thus a better sample of the student's universe of knowledge is possible and, with care in designing the test, partial credit can be allotted even in a machine-scored approach. These conditions could be equally well met using computers, although it would admittedly require careful programming. In addition, the computer has an advantage in that immediate feedback can be provided, further strengthening the reinforcement argument.

Another form of formative assessment is the use of the computer to evaluate student data collected in laboratory exercises. Such

checking of data and calculations is repetitive, prone to error, and not cost effective when done by humans. Computers, on the other hand, excel at this type of task (Harrison and Pitre, 1983, 1988). Programs used in this way are designed to check for realistic values, a range of data, and values clearly outside acceptable limits. When incorrect answers are given, students may be asked to redo their calculations and submit revised figures (May, Murray and Williams, 1985). The programs also may be designed to tentatively accept answers within a certain range, but to suggest that students return to places of potential error and check their work (Harrison and Pitre, 1988).

As part of a project to integrate computer-generated homework into physical science college courses, Milkent and Roth (1989) used computer-generated problems as homework assignments and monitored student progress with computer-generated multiple choice quizzes. They found that the use of the computer-generated homework significantly reduced the effectiveness of ACT scores as predictors of course achievement. Put in other words, as a result of the homework approach, students had greater opportunities for achieving mastery and for minimizing the potential influence of entry level aptitude and prior academic preparation. This was in addition to the teacher advantages of an efficient system for homework management and freedom from bookkeeping procedures.

Incorporation of computers into science instruction often takes the form of microcomputer-based laboratories (MBL). Assessment is frequently a part of such a system. However, in some cases this means simply presenting multiple choice questions by means of the computer screen (Bross, 1986). If immediate feedback is not available, no learning gains may accrue to such computer use. Increased ease of data collection and processing may still make this approach to testing of value to the instructor. A more useful approach might be that described by Browning and Lehman (1988) for identifying student misconceptions in genetics problem solving. Four computer programs were presented and the students' responses were recorded and analyzed for evidence of misconceptions and difficulties in the problem solving process. Three main problem areas were identified: difficulties with computational skills, difficulties in the determination of gametes, and inappropriate application of previous learning to new problems. Evaluation of this type would seem to show considerable promise for remedial instruction and improved student learning.

Collins (1984) conducted a study to determine whether learning would be improved with computerized tests. The students (n=210) were enrolled in a one-semester introductory biology course. Students in the computer section took computer generated tests in addition to the tests taken by students in the other sections. Students taking the computer tests were given immediate feedback on their scores, then told which responses were correct and which were incorrect. In addition, the computer recorded student data on disk, allowing for later analysis by the instructor. Collins concluded that computer testing led to enhanced learning as indicated by higher scores on weekly in-class written tests, the midterm examination, the final examination, and final class marks.

Collins and Earle (1989-90) examined the effects of computer-based learning and computer-administered testing in an introductory biology class. They found that the greatest benefit was attained by those using the computer units in addition to attending regular lectures. Taking weekly computer-administered multiple choice tests also appeared to benefit students of middle and upper ability but not students of lower ability levels. That the use of weekly computer-tests can increase students' scores reinforces a finding of an earlier study (Collins, 1984). Although students benefitted from using either the computer learning units or the computer tests, the use of the two together did not result in even more gain, as might have been expected. Frequency of use of the units appeared to be a factor in that the "frequent" user group achieved a much higher mean score and higher pass rate than did the "infrequent" user group.

The possibility that students were being disadvantaged by taking computer tests instead of written paper forms of the same tests was studied by Fletcher and Collins (1986-87). They found that students' mean scores on the computer-administered test and the written forms of the same test were roughly equivalent, and concluded that the students were not disadvantaged by taking the computer tests. The students indicated that most of them favored the computer-administered tests and cited several major advantages: (1) immediacy of scoring; (2) immediate feedback on incorrect answers; (3) more convenient, straight forward and easy-to-use; and (4) faster than written tests. Two major disadvantages were noted by the students: (1) not being able to review all their responses at the end of the test and make changes; and, (2) not being able to skip questions and come back to answer them later (p. 42).

The converse case was studied by Jackson (1988) who attempted to discover whether a computer could give any significant educational advantage to the pupil. That is, could the computer improve pupil motivation during the test, by giving instant feedback and marking, thus improving understanding and hence give an enhanced score in a future test? (p. 809) The middle school science students who were tested by computer and given immediate feedback scored significantly higher in a later test using the same material than did those students who were tested using the traditional paper and pencil method. An additional gain for the teacher was the ability to conduct further analyses, such as test item analysis, on the computer-recorded student data; such analyses could not be easily carried out without computer administered testing.

Moe and Johnson (1988) investigated students' reactions to a computerized adaptive ability test and examined the practicability of this testing method in the classroom. The students in the study included 161 females and 154 males, fairly evenly divided among grades eight through twelve and including a few college students. The subjects took a computerized version and a printed version of a standardized aptitude test battery and a survey assessing their reactions. Overall reactions to the computerized test were overwhelmingly positive (p. 79). Only 8.6 percent of the students were using a computer for the first time in taking the test. Analysis showed no difference in performance between first time users and those who had had prior experience. More than half the students (51%) reported no difference in the amount of nervousness they felt in taking the computerized version compared to the printed version of the test. Girls were more likely ($p<.05$) to report nervousness than boys, but analysis of variance revealed no significant difference in performance between boys and girls.

The effects of microcomputer-administered diagnostic testing on both student achievement and attitudes were of concern to Waugh (1985). Students in one group were given the unit objectives and responded to a computer-administered diagnostic test consisting of one item per objective. The other group received the objectives and were assigned an out-of-class task of completing an objective specific mini-project. The results showed that microcomputer-administered diagnostic testing could positively influence the immediate achievement of students in science. Evidence did not, however, support the hypothesis that an exposure to diagnostic testing might influence continuing achievement. The findings indicated that the

use of microcomputer-administered diagnostic testing was successful in increasing student achievement in science by an average of six percent with no loss of positive attitude toward school, learning, or science. The evidence further indicated that diagnostic testing might have played a role in arousing student interest in microcomputers.

Student attitudes were also the focus of a study by Knight and Dunkleberger (1977) in a comparison of computer-managed self-paced instruction with teacher-managed group-paced instruction for ninth grade students. The course consisted of large group lectures (31% of the overall time), small group seminars (46% of the time), and laboratory activities (31% of the time). The computer-managed self-paced group and the teacher-managed group-paced students received the same large group lectures and small group seminars. The computer group was allowed to self-pace through the laboratory activities while the teacher-managed group followed a group-pace. The computer served as an assessment and record keeping device for the computer-managed students. The quizzes were four-choice, multiple choice questions and students received immediate feedback after completing each item. Although the differing instructional approaches were applied only during the laboratory component of the course (31%), the positive reaction of the computer-managed self-paced group was sufficiently strong to effect a significant difference in attitudes toward the study of science.

The impact of an emerging technology, interactive videodisc, was studied by Huang and Aloi (1991) in a first year biology course. The interactive video involved 17 menu driven chapters integrating computer text with laser disc images and computer graphics. The students were organized into groups with inter-group competition in answering true/false, multiple choice, and completion questions. The researchers compared, using an unpaired t-test, the proportion of students getting A, B, C, D, F, and W (withdraw) for 11 semesters prior to using interactive video with the proportions during the 5 semesters following its use. They found that the proportion receiving A's increased significantly ($p<.005$) following use of the interactive video. The percentage increases were: A's, 6% before and 18% after; B's, 21% before and 32% after; C's, 20% before and 36% after; D's 10% before and 4% after; F's did not change. Retention of students was also increased. The proportion of W's was 33% before interactive video use and 24% after. Thus, the use of interactive videodisc resulted in increased proportions of success at nearly all levels of achievement.

Interactive videodisc was also used as a tool in assessing science teachers' knowledge of safety regulations in school laboratories for purposes of teacher certification by the Connecticut State Department of Education (Lomask, Jacobson and Hafner, 1992). The program simulates a typical lab activity in a secondary school general science course and shows four student performing a simple lab experiment to identify unknown materials. The IVD assessment includes two stages: stage one deals with safety equipment and storage of chemicals and stage two deals with students' laboratory practices. The examinees are asked to assume the role of the lab teacher by viewing an interactive videodisc simulated classroom. The teachers are then asked to identify safety violations and to suggest preventive or corrective measures. Subjects responses are recorded for later analysis and scoring (p. 1).

There appear to be several advantages to incorporating some form of computer assistance in assessment. Immediate feedback to the students seems to be a consistent factor in increased achievement. Ease of test taking, together with improved record keeping, suggest improved efficiency for both students and teachers. The availability of large test item banks makes possible several intermediate quizzes with apparent achievement gains the result. Such formative evaluation serves both as a diagnostic tool and as a remediation device, indicating where corrections are needed. The data collection capability of computer testing also permits more extensive data analysis, especially in the area of test item analysis, which in turn should yield more reliable and, presumably, more valid assessment. Two cautions must be noted, however. First, the simplicity of devising multiple choice, true/false, matching, and other objective tests can lull the teacher into simply doing a better job of assessing low level recall knowledge. Second, the linear nature of most computer testing does not allow the student to go back and reflect upon a particular item, nor to view the completed test as a whole to check for consistency of responses. The increased improvement and implementation of such emerging technologies as interactive video and hypermedia (Kumar, 1991) show high promise for overcoming both difficulties by providing opportunities for both improved levels of questions and increased flexibility in the testing process.

Alternative Assessments

Assessment attention primarily has been focused, perhaps understandably, on achievement in science. With regard to assessment

of achievement, Shavelson, Carey and Webb (1990) note:

> Unfortunately, in an attempt to create achievement tests in science and other subjects that do not unduly favor one or another of the nearly infinite number of curricula in our country, the current technology produces tests that emphasize recall of facts and performance of isolated skills but tend not to measure students' conceptual understanding and problem-solving skills. Consequently, the current technology works against what many people value as education (p. 697).

Paper and pencil. testing might suffice for determining how much, and what, students know if science programs were designed only to acquire content. As more problem-solving and process skills are incorporated into the science program, different forms of assessment are required to measure these learning outcomes (Meng and Doran, 1990). A variety of assessment methods are suggested by Meng and Doran. In addition to paper and pencil tests, there are practical tests, observations, discussions (or interviews), practical tests in which students manipulate materials, and projects and written work in which the students can demonstrate investigative research or construct something. It is necessary, of course, that the type of assessment be appropriate for the learning outcomes being measured. Although paper and pencil tests are currently regarded with suspicion, they are adequate for measuring knowledge of certain kinds of content, such as facts and terminology. They do not, however, effectively measure whether or not students can apply their knowledge. Application of knowledge involving problem-solving and process skills are better assessed by means of practical tests or observations of student performance.

In the belief that assessment techniques beyond the traditional paper and pencil tests were needed, Doran, Boorman, Chan and Hejaily (1992a) conducted a study to develop and validate instruments to assess the level of laboratory skills possessed by students completing the high school science courses (biology, chemistry, and physics). More than 1000 students from 35 schools participated in the study. It was decided that the "whole investigation" format of laboratory practical testing would be the model used and, in each of the science areas, tests were developed around six laboratory tasks (Doran, Boorman, Chan and Hejaily, 1992b). The model used was one developed by Tamir, Lunetta and colleagues and is composed of

stages that are congruent with the prelab/lab/postlab format of many inquiry-oriented science programs (Lunetta, Hofstein and Giddings, 1981; Uri and Hofstein, 1982). The three stages used were planning, performing, and reasoning. A two-part test was designed to account for the different kinds of skills needed in the planning/design stage and those required in the other two stages. The scoring system developed was general in the sense that it was applicable across each of the tasks in biology, chemistry, and physics. The scoring was subjective in that the test booklets were scored by raters. Thus, test reliability, inter-rater agreement, and correlation between raters had to be determined. Analyses of the data indicated that the reliabilities and correlations were sufficient to warrant further development and investigation.

A national assessment program in the United Kingdom is described by Burstall (1986). The target audience for science was 11-, 13-, and 15-year-olds. The overall assessment included both an oral and a practical component. The oral test was designed to determine the student's ability to communicate effectively and deals with general rather than science-specific information. The science practical test included both a script to guide the examiner's verbal assessment of student understanding and a checklist to record observations of the student's performance. The major advantages of the practical assessment were found to include: (1) All question are oral; poor readers are not penalized. (2) Students may ask for clarification of the test. (3) The pace and extent of the testing can be adjusted to suit the student. (4) Students have the opportunity to retract or amend an answer. (5) Assessors are permitted to prompt in order to direct students toward an appropriate strategy. (6) Assessors are able to observe the method and problem-solving strategies of the student (p. 18).

Keys (1992) describes a study to examine the converse of the procedure described above in that the students' written laboratory reports were analyzed for evidence of conceptual and procedural understandings in science. The eighth grade students involved did two laboratory activities, one on the inclined plane and one on the action of levers. The students had been previously trained to write laboratory reports in full sentences using a specified structure including, problem, hypothesis, materials, procedure, data/observations, conclusion and discussion sections (p. 4). Questioning prompts developed by the researcher were included in the report-writing stage to encourage the students to write more about the

procedures and the concept involved in the activities. The laboratory reports were analyzed for conceptual and procedural understandings, and a scoring guide was developed using propositional analysis and rating scales. No reliability was determined for the scoring guide, this is intended to be further investigated and refined. Qualitative analyses of the students' reports revealed many naive conceptions and instances where students used the data they had collected to support their naive conceptions. There was a tendency for students who had collected reasonable data to draw conclusion which the data did not support. In some cases, students drew conclusions in the opposite direction from that which could be inferred from the data. Others had unreasonable data, but still arrived at correct conclusions. Another tendency was to oversimplify the results in an attempt to make the drawing of conclusions more manageable. While there are still some evident problems to be solved with this approach to assessment, it does provide another option to be further pursued.

Many other options exist, of course. The March, 1992, issue of *Science Scope* is devoted to alternative forms of assessment. Similarly, a chapter in *Science Assessment in the Service of Reform* also includes descriptions of assessment alternatives (Kulm and Malcom, 1991). Articles are included on performance-based assessment, portfolio assessment, group assessment (involving team approaches), concept mapping, scoring rubrics (techniques), dynamic assessment, and assessment for individual differences. While these approaches, for the most part, represent still-emerging options with validity and reliability still to be established, they nonetheless offer ways to broaden assessment and improve the prospects for gaining a more complete picture of student understanding.

Alternative approaches to assessment offer both encouragement and opportunities. The encouragement stems from the possibilities we gain to develop a more complete perspective of student knowledge and understanding. The opportunities derive from the work still to be done in determining the validity and reliability for most of the approaches. Most of these alternatives are more time consuming than is traditional testing. With the increased availability and flexibility of computer and related technology, the cost in time and effort can be sharply reduced. The additional data collecting, storing and analysis capability made possible with the computer makes alternative forms of assessment increasingly attractive.

References

Aesche, Darryl W. and Graham R. Parslow. (1988). "Use of Questions from the Medical Biochemistry Question Bank with the 'Q' Instruction Package." Biochemical Education, 16(1): 24-28, January.

Aldridge, Bill G. (1989a). "Essential Changes in Secondary School Science: Scope, Sequence, and Coordination." NSTA REPORT. Washington, DC: National Science Teachers Association.

Aldridge, Bill G. (1992). "Project on Scope, Sequence, and Coordination: A New Synthesis for Improving Science Education." Journal of Science Education and Technology, 1(1): 13-21.

Aldridge, Bill G. (1989b). "Scope, Sequence, and Coordination of Secondary Science. A Rationale." Washington, DC: National Science Teachers Association.

"Alternative Assessment and Technology." (1992). News from the Center for Children and Technology. 1(3): 1-5, March. Bank Street College of Education, NY.

Blosser, Patricia E. (1984). "Attitude Research in Science Education." Information Bulletin No. 1. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education, The Ohio State University.

Bross, Thomas R. (1986). "The Microcomputer-Based Science Laboratory." Journal of Computers in Mathematics and Science Teaching, 5(3): 16-18, Spring.

Browning, Mark E. and James D. Lehman. (1988). "Identification of Student Misconceptions in Genetics Problem Solving via Computer Program." Journal of Research in Science Teaching, 25(9): 747-761, December.

Burstall, Clare. (1986). "Innovative Forms of Assessment: A United Kingdom Approach." Educational Measurement: Issues and Practice, 5(1): 17-22, Spring.

Collins, Angelo. (1992). "Portfolios: Questions for Design." Science Scope, 15(6): 25-27.

Collins, Michael A.J. (1984). "Improved Learning With Computerized Tests." The American Biology Teacher, 46(3): 188-190.

Collins, M.A.J. and P. Earle. (1989-90). "A Comparison of the Effects of Computer-Administered Testing, Computer-Based Learning and Lecture Approaches on Learning in an Introductory Biology Course." Journal of Computers in Mathematics and Science Teaching, 9(2): 77-84, Winter.

Dawson, George. (1987). "10,000 Science Questions by Modem." The Science Teacher, 54(3): 41-44, March.

Doran, Rodney L., Joan Boorman, Fred Chan and Nicholas Hejaily. (1992a). "Alternative Assessment of High School Laboratory Skills." Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA, March 21-25.

Doran, Rodney L., Joan Boorman, Alfred Chan and Nicholas Hejaily. (1992b). "Successful Laboratory Assessment." The Science Teacher, 59(4): 22-27, April.

Dunkleberger, Gary E. (1980). "Applying 20/20 Hindsight to Self-Pacing." The Science Teacher. 47(8): 32-34, November.

Fletcher, P. and M.A.J. Collins. (1986-87). "Computer-Administered Versus Written Tests--Advantages and Disadvantages." Journal of Computers in Mathematics and Science Teaching, 6(2): 38-43, Winter.

Gardner, P.L. (1975). "Attitudes to Science: A Review." Studies in Science Education, 2: 1-41.

Gauld, C.F. and A.A. Hukins. (1980). "Scientific Attitudes: A Review." Studies in Science Education, 7: 129-161.

Germann, Paul J. (1988). "Development of the Attitude Toward Science in School Assessment and its Use to Investigate the Relationship Between Science Achievement and Attitude Toward Science in School." Journal of Research in Science Teaching, 25(8): 689-703, November.

Harrison, David and John M. Pitre. (1983). "Computerized Lab Test on Error Analysis." The Physics Teacher, 21(9): 588-589, December.

Harrison, David and John M. Pitre. (1988). "Computerized Pre-checking of Laboratory Data." The Physics Teacher, 26(3): 156, March.

Heikenkinen, Henry and Gary E. Dunkleberger. (1985). "On Disk With Mastery Learning." The Science Teacher, 52(7): 26-28, October.

Huang, Samuel D. and Jane Aloi. (1991). "The Impact of Using Interactive Video in Teaching General Biology." American Biology Teacher, 53(5): 281-284, May.

Jackson, Bryan. (1988). "A Comparison Between Computer-Based and Traditional Assessment Tests, and Their Effects on Pupil Learning and Scoring.", School Science Review, 69(249): 809-815, June.

Keys, Carolyn. (1992). "Evaluating Students' Written Reports for Evidence of Conceptual and Procedural Understandings in Science." Poster session presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA, March 21-25.

Knight, Carlton W. II. and Gary E. Dunkleberger. (1977). "The Influence of Computer-Managed Self-Paced Instruction on Science Attitudes of Students." Journal of Research in Science Teaching, 14(6): 551-555, November.

Krynowsky, Bernie. (1988). "Problems in Assessing Student Attitude in Science Education: A Partial Solution." Science Education, 72(5): 575-584.

Kulm, Gerald and Shirley M. Malcolm, Eds. (1991). Science Assessment in the Service of Reform. Washington, DC: American Association for the Advancement of Science.

Kumar, David D. (1991). "A Note on 'Hyperequation'." An occasional paper. National Center for Science Teaching and Learning, The Ohio State University, Columbus, September.

Leuba, Richard J. (1986). "Machine-Scored Testing, Part I: Purposes, Principles, and Practices." Engineering Education, 77(2): 89-95.

Leuba, Richard J. (1986). "Machine-Scored Testing, Part II: Creativity and Item Analysis." Engineering Education, 77(3): 181-186.

Lomask, Michal, Larry S. Johnson and Laurin P. Hafner. (1992). "Interactive Videodisc as a Tool for Assessing Science Teachers' Knowledge of Safety Regulations in School Laboratories." Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA, March 21-25.

Lunetta, Vincent, Avi Hofstein and Geoffrey Giddings. (1981). "Evaluating Science Laboratory Skills." The Science Teacher, 48(1): 22-25, January.

May, Peter M., Kevin Murray and David R. Williams. (1985). "CAPE: A Computer Program to Assist with Practical Assessment." Journal of Chemical Education, 62(4): 310-312, April.

Mayer, Victor J. and James M. Richmond. (1982). An overview of assessment instruments in science. Science Education, 66(1): 49-66.

Mayer, Victor J. (1992). "Earth Systems Education." Earth Systems Education Program, The Ohio State University, Columbus, Ohio.

Mayer, Victor J. (1991). "The Framework for Earth Systems Education." Science Activities, 28(1): 8-9.

Meng, Elizabeth and Rodney L. Doran. (1990). "What Research Says . . . About Appropriate Methods of Assessment." The Science Teacher, 28(1): 42-45, September.

Milkent, Marlene M. and Wolff-Michael Roth. (1989). "Enhancing Student Achievement Through Computer-Generated Homework." Journal of Research in Science Teaching, 26(7): 567-573, October.

Moore, Richard W. and Frank X. Sutman. (1970). "The Development, Field Test and Validation of an Inventory of Scientific Attitudes." Journal of Research in Science Teaching, 7(2): 85-94, June.

Munby, Hugh. (1983). An Investigation Into the Measurement of Attitudes in Science Education. Columbus, OH: SMEAC Information Reference Center, The Ohio State University. (ERIC Document Reproduction Service No. ED 237 347).

Munby, Hugh, Richard Kitto and Robert Wilson. (1976). "Validating Constructs in Science Education Research: The Construct 'View of Science'." Science Education, 60(3): 313-321.

Moe, Kim C. and Marilyn F. Johnson (1988). "Participants' Reactions to Computerized Testing." Journal of Educational Computing Research. 4(1): 79-86

Nachmias, Rafi and Marcia C. Linn. (1987). "Evaluations of Science Laboratory Data: The Role of Computer-Presented Information." Journal of Research in Science Teaching, 24(5): 491-506, May.

Omerod, M.D. and D. Duckworth. (1975) Pupils' Attitudes to Science: A Review of Research. Windsor, England: NFER Publishing Company.

Schibeci, R.A. (1984). "Attitudes to Science: An Update." Studies in Science Education, 11: 26-59.

Science for All Americans: Summary. (1989). Washington, DC: American Association for the Advancement of Science.

Shavelson, Richard J., Neil B. Carey and Noreen M. Webb. (1990). "Indicators of Science Achievement: Options for a Powerful Policy Instrument." Phi Delta Kappan, 71(9): 692-697, May.

Update: Project 2061. Education for a Changing Future. (1992). Washington, DC: American Association for the Advancement of Science.

Uri, Ganiel and Avi Hofstein, (1982). "Objective and Continuous Assessment of Student Performance in the Physics Laboratory." Science Education, 66(4): 581-591.

Vogel, George. (1985). "A Microcomputer-Based System for Filing Test Questions and Assembling Examinations." Journal of Chemical Education, 62(11): 1024-1027, November.

Ward, Thomas J., Simon R. Hooper and Kathleen M. Hannafin. (1989). "The Effect of Computerized Tests on the Performance and Attitudes of College Students." Journal of Educational Computing Research, 5(3): 327-333.

Waugh, Michael Leonard. (1985). "Effects of Microcomputer- Administered Diagnostic Testing on Immediate and Continuing Science Achievement and Attitudes." Journal of Research in Science Teaching, 22(9): 793-805, December.

Waugh, Michael L. and Dawn Currier. (1986). "Computer-Based Education: What We Need to Know." Journal of Computers in Mathematics and Science Teaching, 5(3): 13-15, Spring.